Flujo de limpieza de datos

(eliminar==1), estado (8 o 2)

Accesos AWS: AWS_ACCESS_KEY_ID AWS_SECRET_ACCESS_KEY AWS_DEFAULT_REGION Archivos: Instituciones: s3://conecta-ideas-chile-dumps/database/instituciones.csv.gz Cursos: s3://conecta-ideas-chile-dumps/database/cursos.csv.gz Usuarios por curso: s3://conecta-ideas-chile-dumps/database/usuarios cursos.csv.gz Usuarios: s3://conecta-ideas-chile-dumps/database/usuarios.csv.gz Ejercicios: s3://conecta-ideas-chile-dumps/database/ejercicios realizados.csv.gz Renombrar Guardar archivo: Instituciones columnas, ajustar instituciones *.csv fechas Subset: institucion id, nombre_inst Renombrar Subset: columnas, filtrar institucion_id, inner_join by Cursos grados 4-6, curso_id, grado_curso, institucion_id ajustar fechas seccion curso Filtrar rol = alumnos Subset: Usuarios por usuario_id, curso_id, ajustar fechas, curso añadir columnas creacion, fecha, (creacion, fecha) date_updated Eliminar registros anteriores a "2021-04inner_join by Renombrar Merge by Usuarios columnas usuario id 01", eliminar columnas curso id (creacion, date updated) Éliminar alumnos prueba,` Guardar archivo: remover tildes en seccion curso, eliminar base_sql_*.csv registros con "@" y ".0" Remover pruebas Guardar archivo: de Grade base_sql_singrade_*.csv Eliminar columna Guardar archivo: **Ejercicios** (id), renombrar ejercicios_completa_*.csv columnas Merge by curso id, Guardar archivo: usuario_id ejercicios_realizados_*.csv Subset: Ajustar fechas, creación de Guardar archivo: No tomar en cuenta las columnas (minutes, inicio1, ejercicios_realizados_ pruebas de grade limpio_*.csv termino1)